The future of Earth system modelling

Peter Dueben

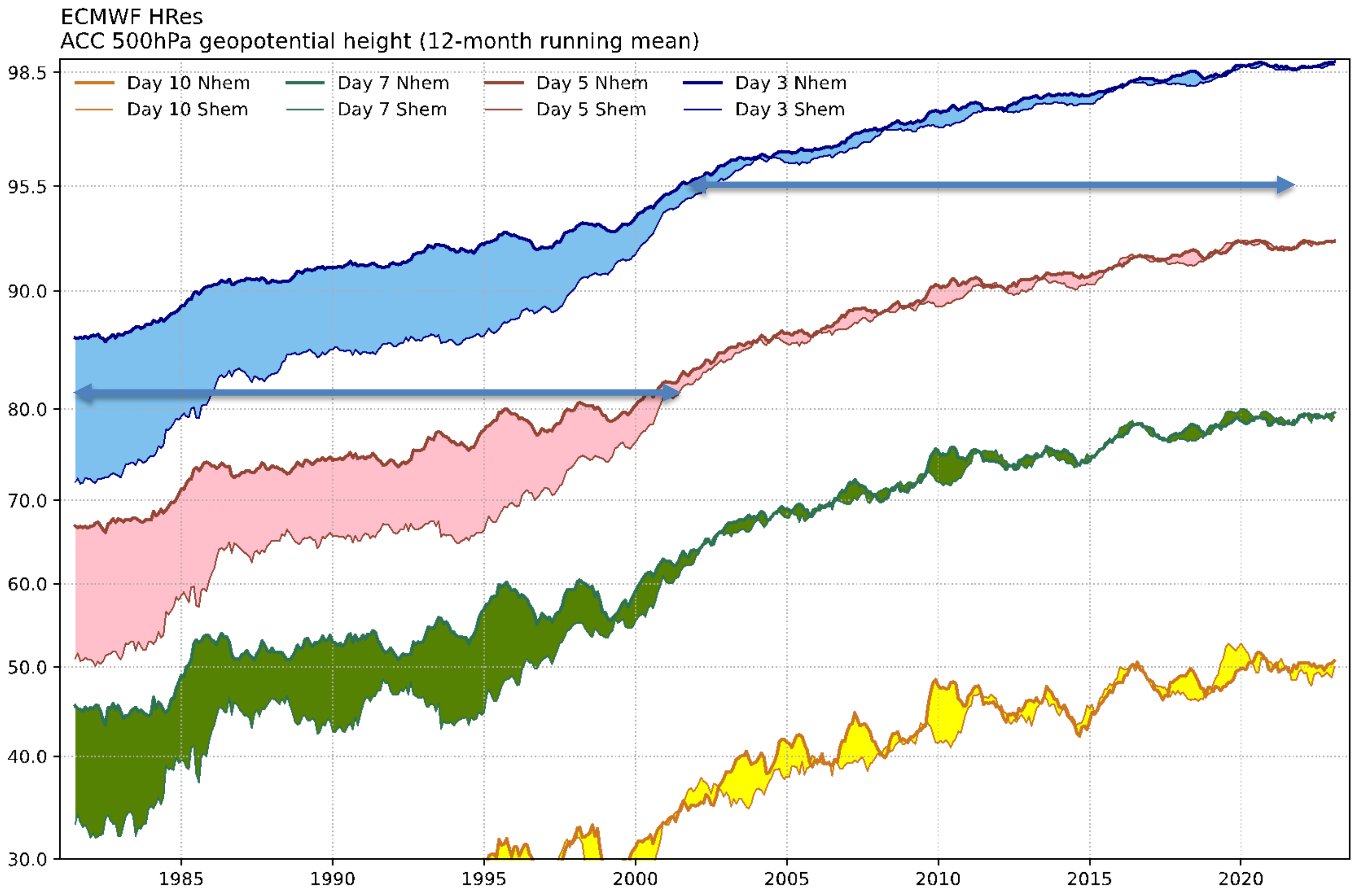
Head of the Earth System Modelling Section

The strength of a common goal

UNIVERSITY OF COLOGNE

The MAELSTROM, ESiWACE and WeatherGenerator projects have received funding from Horizon Europe and the EuroHPC-Joint Undertaking under grant agreement No 955513, 101093054. and 101187947.

The quiet revolution of numerical weather prediction



nature

Explore content | About the journal | Publish with us | Subscribe

nature > review articles > article

Published: 02 September 2015

The quiet revolution of numerical weather prediction

Peter Bauer, Alan Thorpe & Gilbert Brunet

Nature **525**, 47–55 (2015) | Cite this article

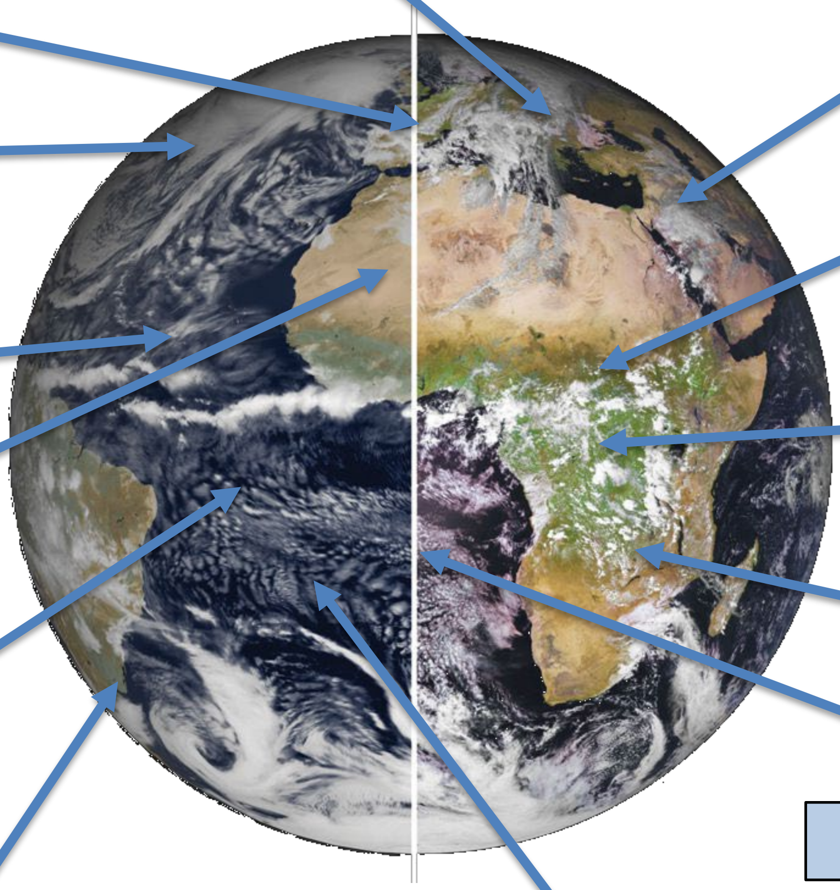
Abstract

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

Key points:

- Predictability improved at roughly one forecast day per decade.
- Improvements via observations, resolution (via HPC power), data assimilation algorithms and improvements of the model.

The physical model of the future – Plenty of great work ahead for modelling



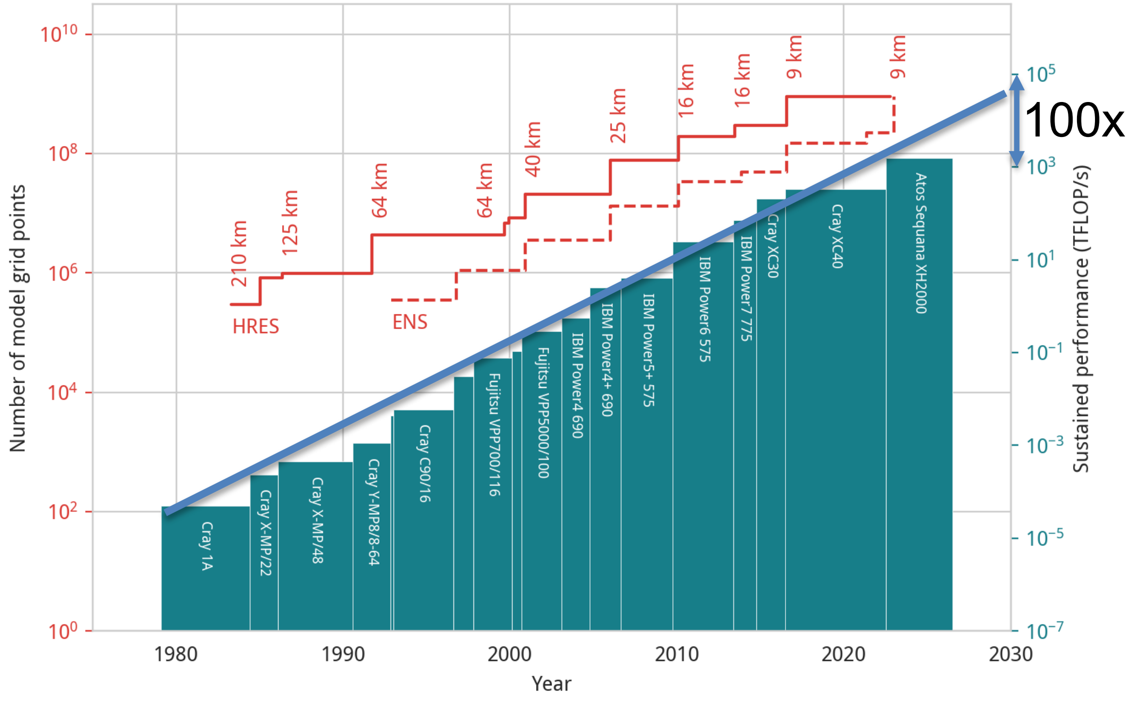
All great, but there will be no sudden jumps in forecast scores.

The quiet revolution of numerical weather prediction

HPC: GPUs are still a nightmare to use, too expensive and Moore's law is dead.
Scientific computing plays no role for future HPC hardware.

Km-scale models are now becoming "standard":
Improved precipitation, improved tropical cyclones, improved 2mT, improved topography representation…
But gains for global forecast scores are limited.

Observations and data assimilation methodology:
Not for me to comment.



ECMWF

The ~~quiet~~ ML revolution of numerical weather prediction

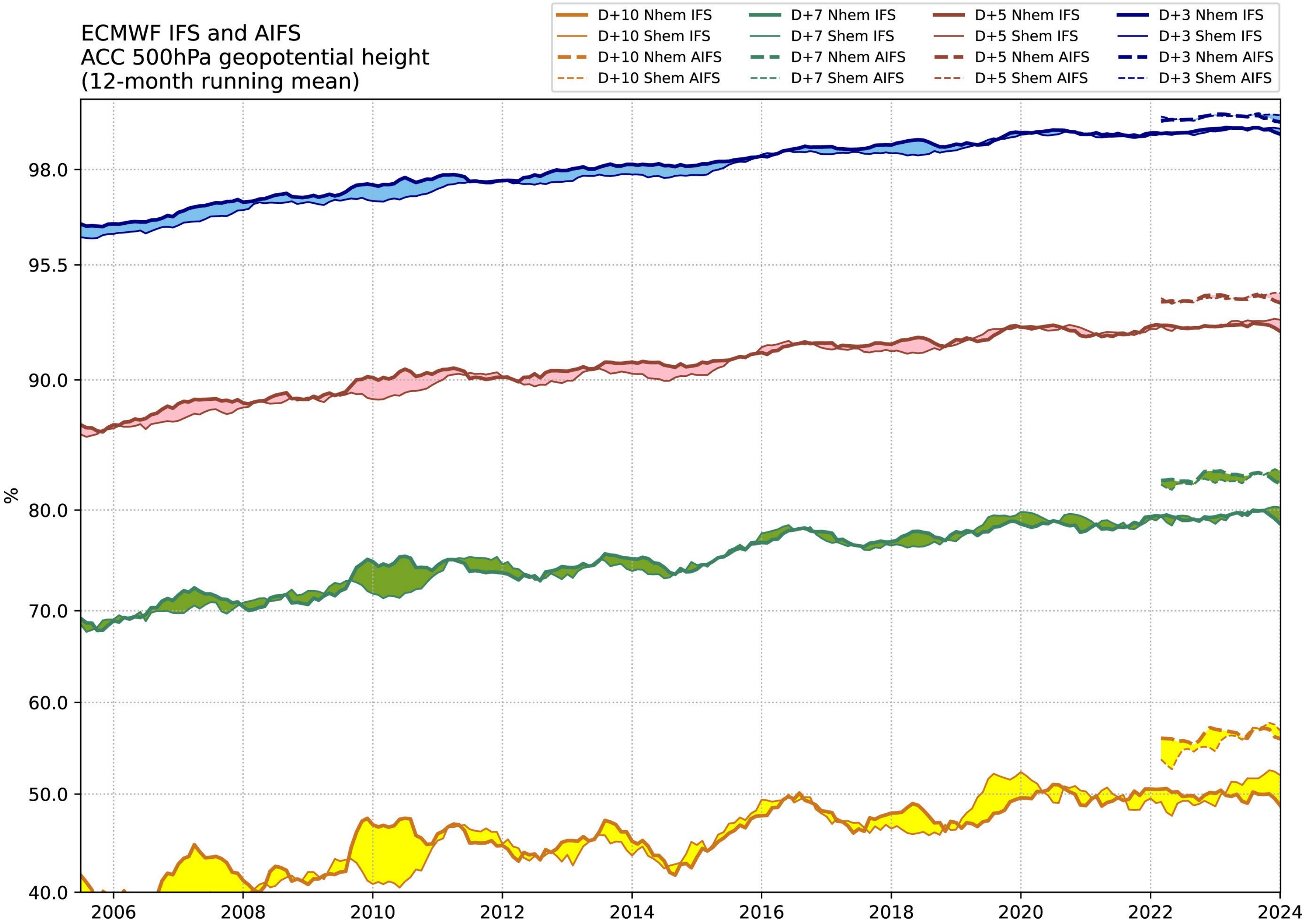
The quiet revolution is dead…

…long live the machine learning revolution.

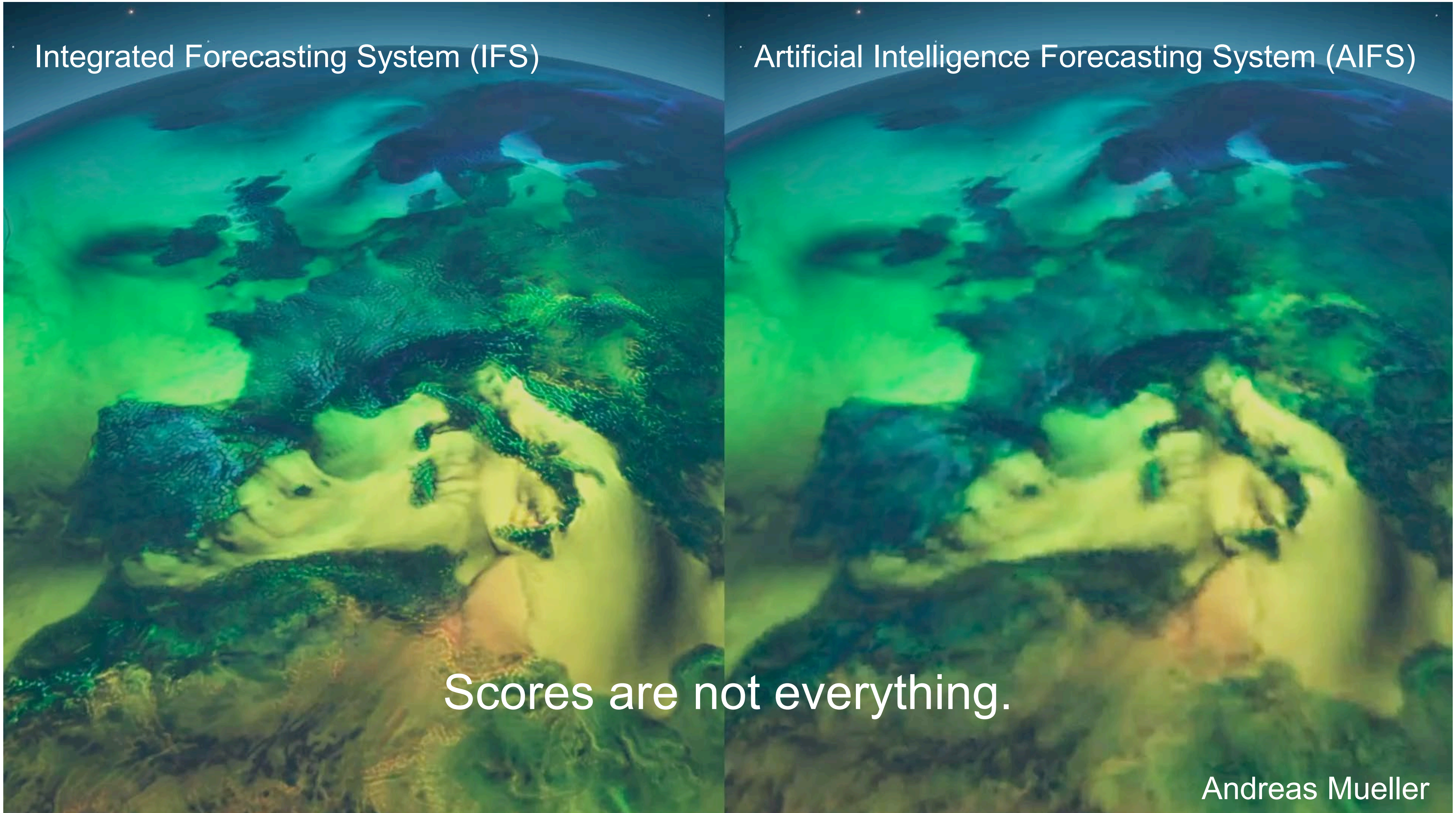
AIFS is years ahead of IFS in terms of forecast scores for deterministic and ensemble predictions.

If you want to have competitive scores with IFS, you need to nudge it to AIFS.

And we can trust weather predictions of machine-learned models.



The machine learning revolution



The machine learning revolution

It is so difficult to see what machine learning will bring in the next five years…

Really?

Universal approximation theorem:
Machine learning can learn everything if you have enough data and compute.

Easy questions:
Do we have the right software? – Yes.
Do we have the right algorithms? – Yes.
Do we have enough compute? – Not at home, but at EuroHPC. → Yes.

Fundamental questions:
How many models will we need?
Do we have enough data?

How many models do we need?

Weather Generator

Task specific training for languages:

Input: English text → Output: Spanish text

Inference for task specific training for languages:

Input: English text → Output: Spanish text → Translate English to Spanish

Large Language Models (LLMs):

Input: All text available → Output: Gapfill all text available

Training on all text is helping the neural network to learn a general model of natural language.

LLMs inference for languages:

Input: I want a recipe for a vegan lasagne in Spanish → Output: Spanish recipe

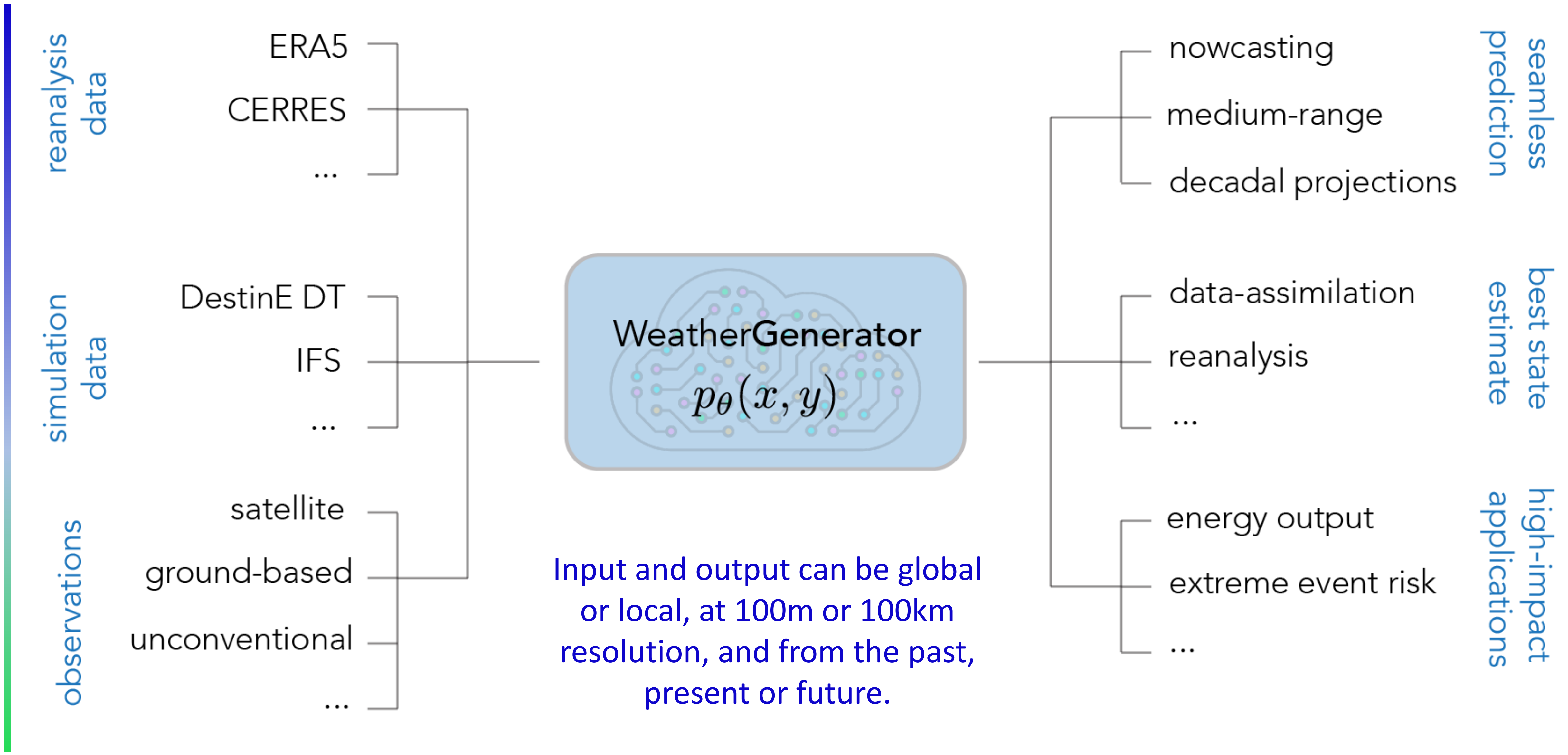
Input: Translate this English text into Chinese → Output: Chinese text

Input: How would Shakespeare have written this letter? → Output: Letter in Tudor style

…

Fine-tuning and model distillation can help to calibrate LLMs for a specific application to make them more accurate, faster and cheaper.

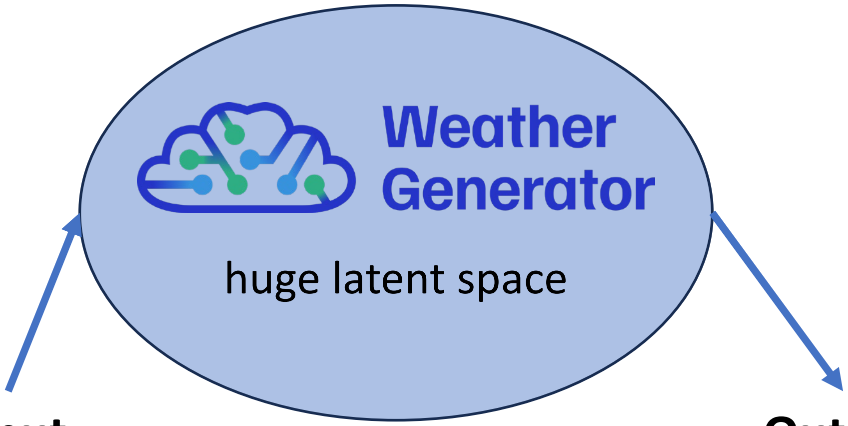
Let's use the WeatherGenerator



Input and output can be global or local, at 100m or 100km resolution, and from the past, present or future.

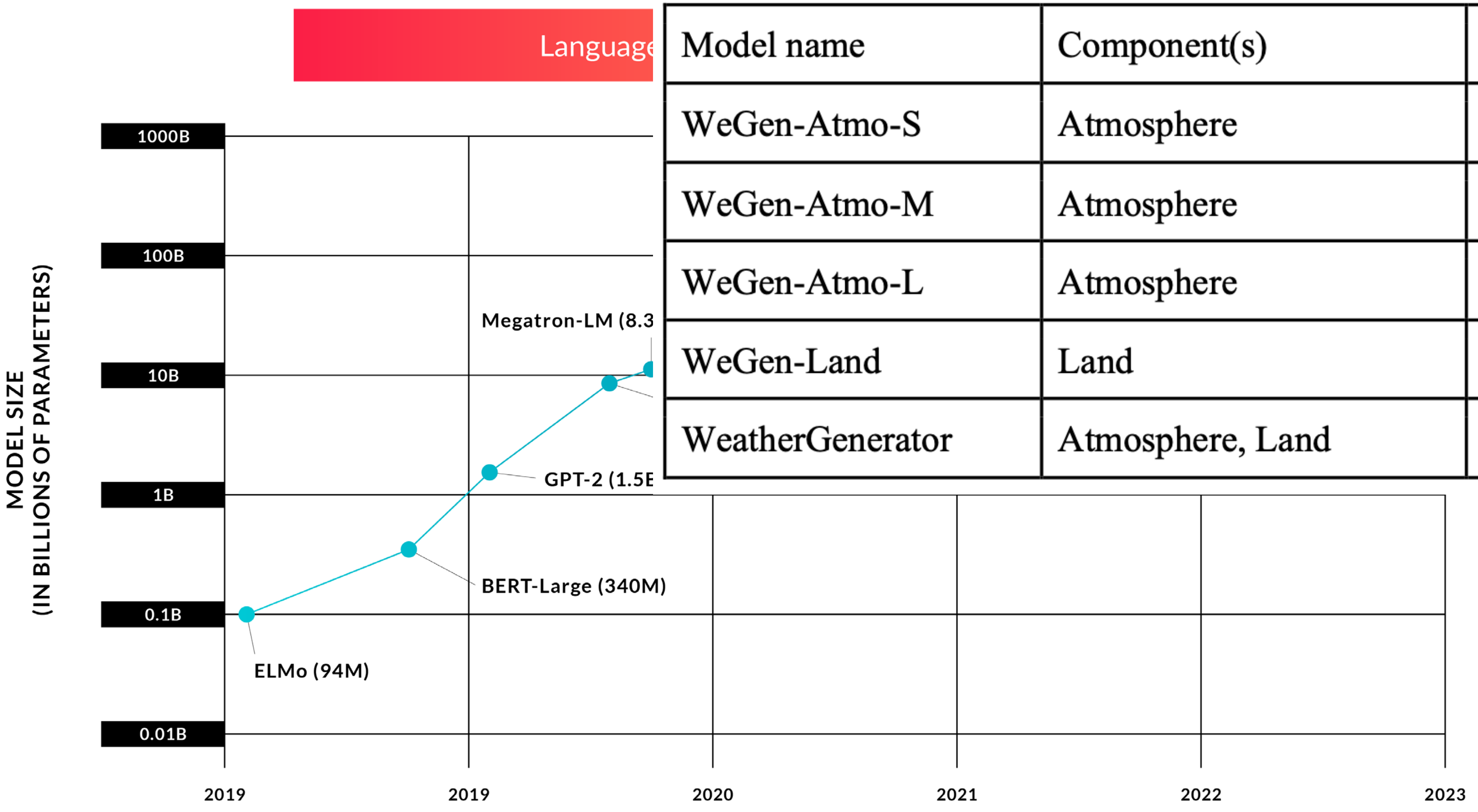
Figure and concept from Christian Lessig

How to use the WeatherGenerator?



Input		Output		Application
Satellite and synoptic observation	→	ERA (35km) and CERRA (5km)	→	Data assimilation
IFS initial conditions (9km)	→	IFS state in the future (9km)	→	Global weather forecast
IFS forecast (9km)	→	OPERA (2km)	→	Model downscaling
DestinE Extremes Twin (4km)	→	synoptic observations	→	Post-processing
Sentinel-1	→	AROME (2km)	→	Observation operator
DestinE Climate Twin (4km)	→	ECOSTRESS (100m)	→	Vegetational water stress predictor
…				

Why do we need a WeatherGenerator?

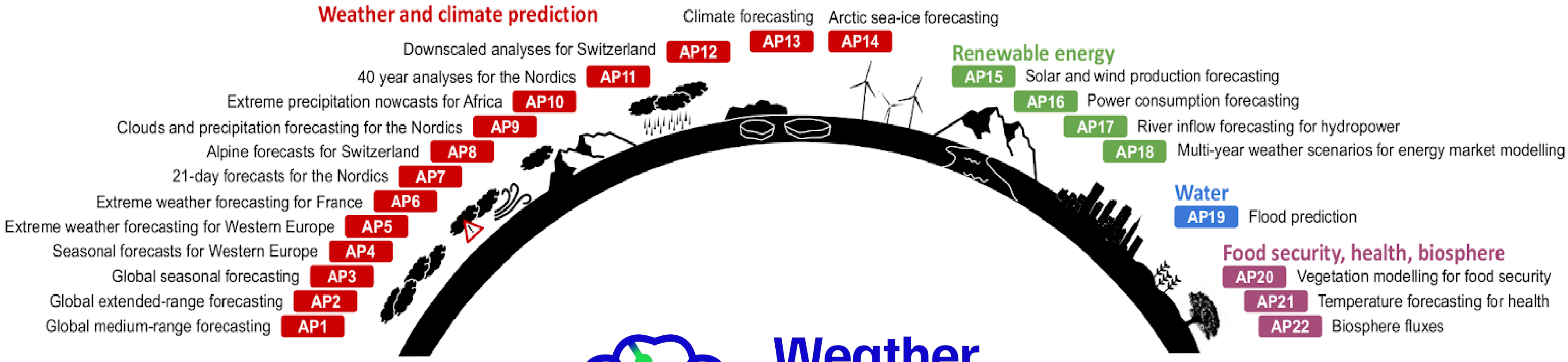


Model name	Component(s)	Model size
WeGen-Atmo-S	Atmosphere	4-6 billion
WeGen-Atmo-M	Atmosphere	10-20 billion
WeGen-Atmo-L	Atmosphere	50-100 billion
WeGen-Land	Land	100 million
WeatherGenerator	Atmosphere, Land	4-6 billion/50-100 billion

Source: https://twosigmaventures.com/

Lang et al. 2024 ensemble AIFS

The aim of the WeatherGenerator



Weather Generator

Aim: This project will build the machine-learned WeatherGenerator – the world's best generative Foundation Model of the Earth system – that will serve as a Digital Twin in Destination Earth (DestinE).

But there is no guarantee that one foundation model will eventually be better when compared to task specific models… ;-(

https://www.ecmwf.int/en/about/media-centre/news/2024/weathergenerator-project-aims-recast-machine-learning-earth-system

How many models do we need?

The gene pool of deep-learning seems to decrease and not increase in recent months.

Representing and understanding individual model components is still useful.

Domain expertise is still essential to design a machine-learned forecast system.

The Earth system is still way too complex to understand the behavior of many task-specific models.

We will not trust models that can only do one thing but behave strangely in others.

Seamlessness is still useful.

To train a leading model will become more and more expensive (for applications with enough data).

My prediction:
We will only need one or two machine-learned Earth system models in the future.
But the "model" is defined by the latent space and not the timestepping scheme.

Do we have enough data?

For parametrization emulation – Yes, but pointless.

For global medium-range weather predictions – Yes, we have ERA.

For longer predictability horizons – The more data, the longer the horizon – Butterfly effect?

For seasonal predictions – Depends on the question and the approach.

For global data assimilation – To be seen.

For land modelling – Yes.

For sea-ice modelling – Yes.

For the ocean – Yes for the surface, no for the deep ocean.

For global atmospheric chemistry – Yes to emulate, no to become better than physical models.

For climate simulations to measure climate sensitivity – No.

For studies of climate tipping points – No.

For extreme weather events in a future climate – Yes, in combination with physical models.

For local down-scaling – Yes for 2mT. Partly for precipitation. Data of physical models for 3D.

For prediction of grey swan events – Yes, when using data of physical models.

Do we have enough data?

In general: What do we do if we do not have enough data?

Use the physical model to generate the data.

What type of physical models do we need?

The best possible configuration with the best possible resolution with the best possible realism to generate the data.

→ Km-scale modelling and Destination Earth

How do we need to adjust physical models?

Physical models will be replaced by machine learning models in many applications
→ efforts in development will reduce.

But this is not the only threat for Earth system models...
If we do not manage to make our models more flexible and easier to use, they will die.
There will be no compiler/software/hardware/staff available for them.

Physical models of the future need to:

- support global, regional city-scale and single column modelling across all timescales,
- address the needs of climate services and DestinE,
- be as realistic as possible in all details,
- be portable and efficient on CPU and GPU hardware,
- be written with a high-level of abstraction that is as similar as possible across all modules,
- be DA ready and differentiable,
- be able to link to machine learning tools for hybrid modelling.

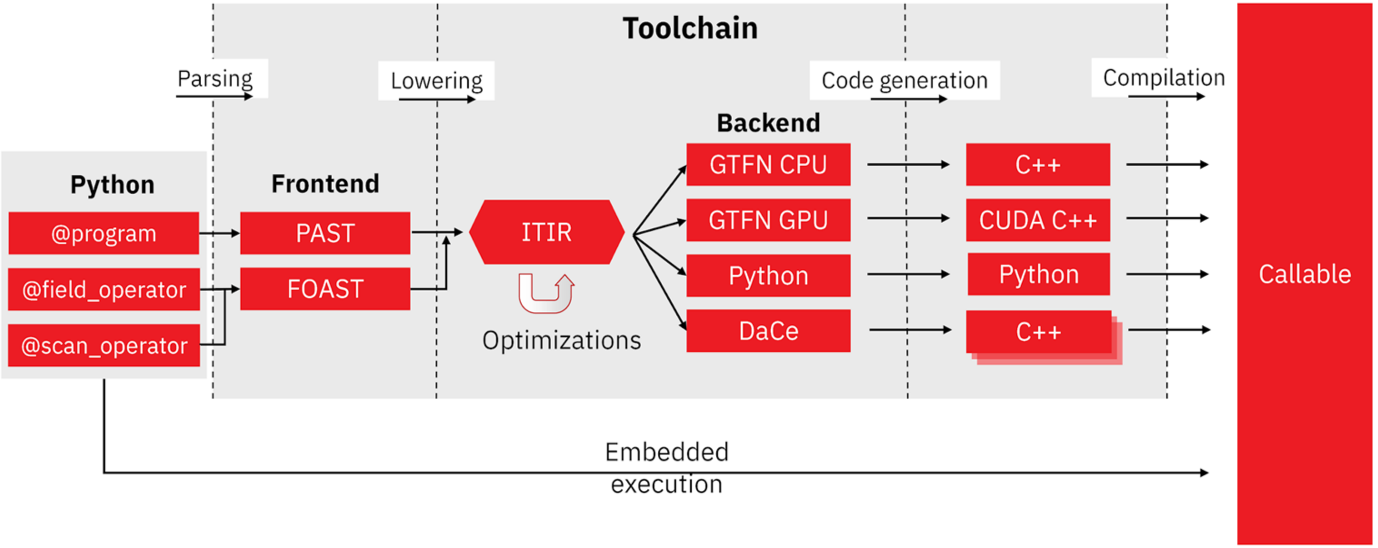
This will need very significant efforts leading to even more centralisation.

The future of the physical models

PMAP – formerly known as the Finite Volume Model (FVM) – as new dynamical core at ECMWF

Python finite volume model with GT4Py backend:
- Local formulation as a finite volume model
- Runs globally and locally
- Scalable
- High-level code base due to use of GT4Py domain specific language (DSL) → portability
- Potentially running in a Jupyter notebook
- Easy to couple with machine learned models
- Potentially transferable to JAX → differentiable

Where is the catch?
How do we bring the model back into the IFS forecast system?



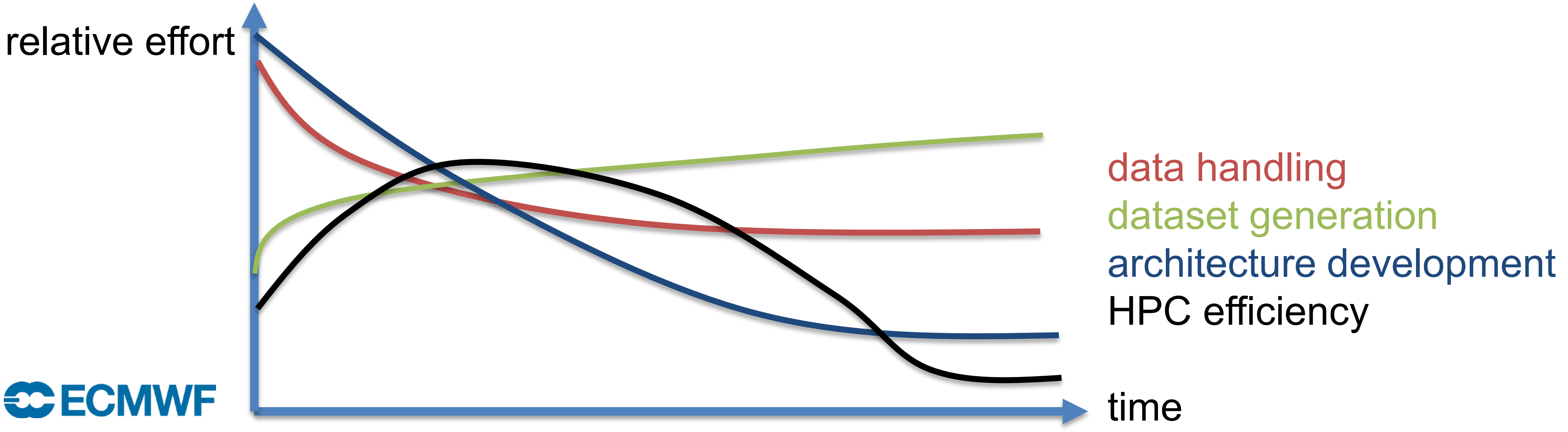
ECMWF

The future of machine-learned models

To support machine-learning in Earth system modelling, we need to build general open-source infrastructure:

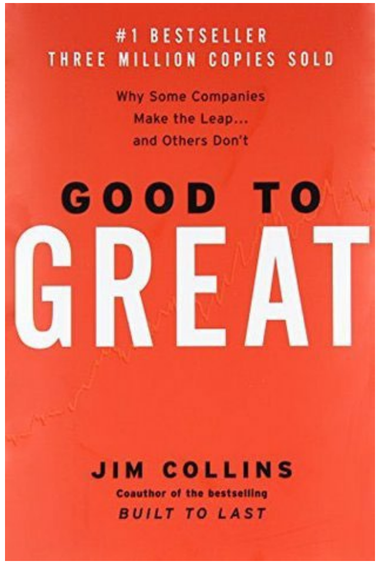
- Anemoi supporting local and global weather predictions
- AIFS as NWP model
- AIFS as full Earth system model including CO2 and ocean, land, sea-ice and wave components
- WeatherGenerator
- Benchmark datasets
- Training datasets for global and local weather of the past, present and future

...and the community will pick it up, supported by programmes such as Copernicus and DestinE



ECMWF

The future of Earth system modelling

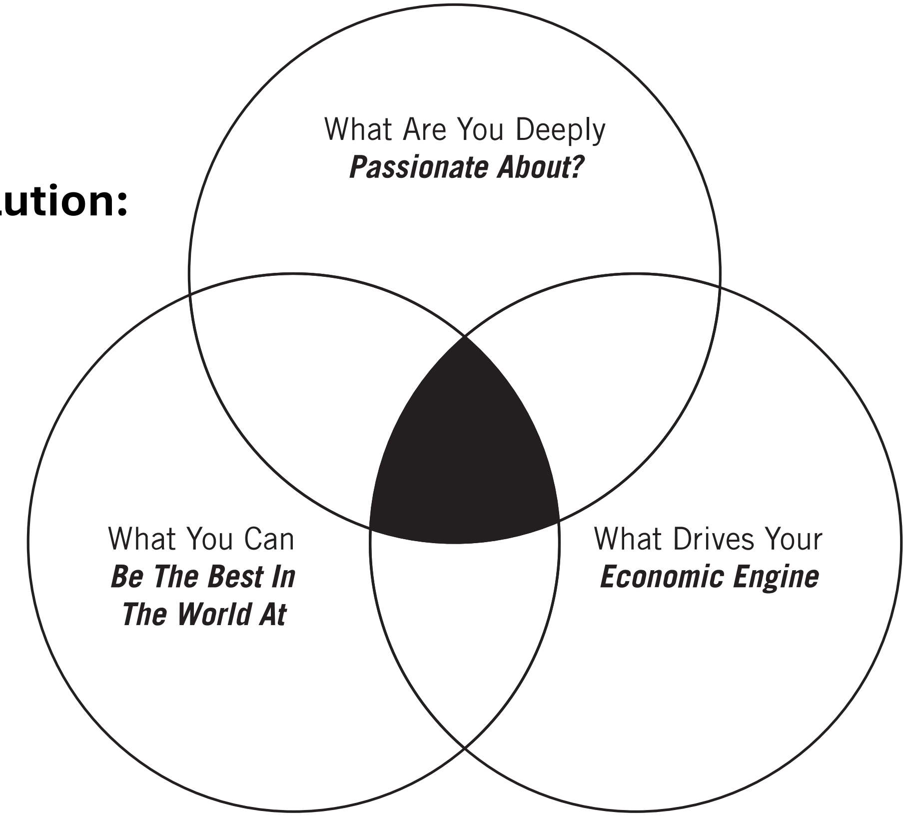


The Hedgehog Concept by Jim Collins: A simple, crystalline concept that flows from deep understanding about the intersection of three circles: 1) what you are deeply passionate about, 2) what you can be the best in the world at, and 3) what best drives your economic or resource engine. Transformations from good to great come about by a series of good decisions made consistently with a Hedgehog Concept, supremely well executed, accumulating one upon another, over a long period of time.

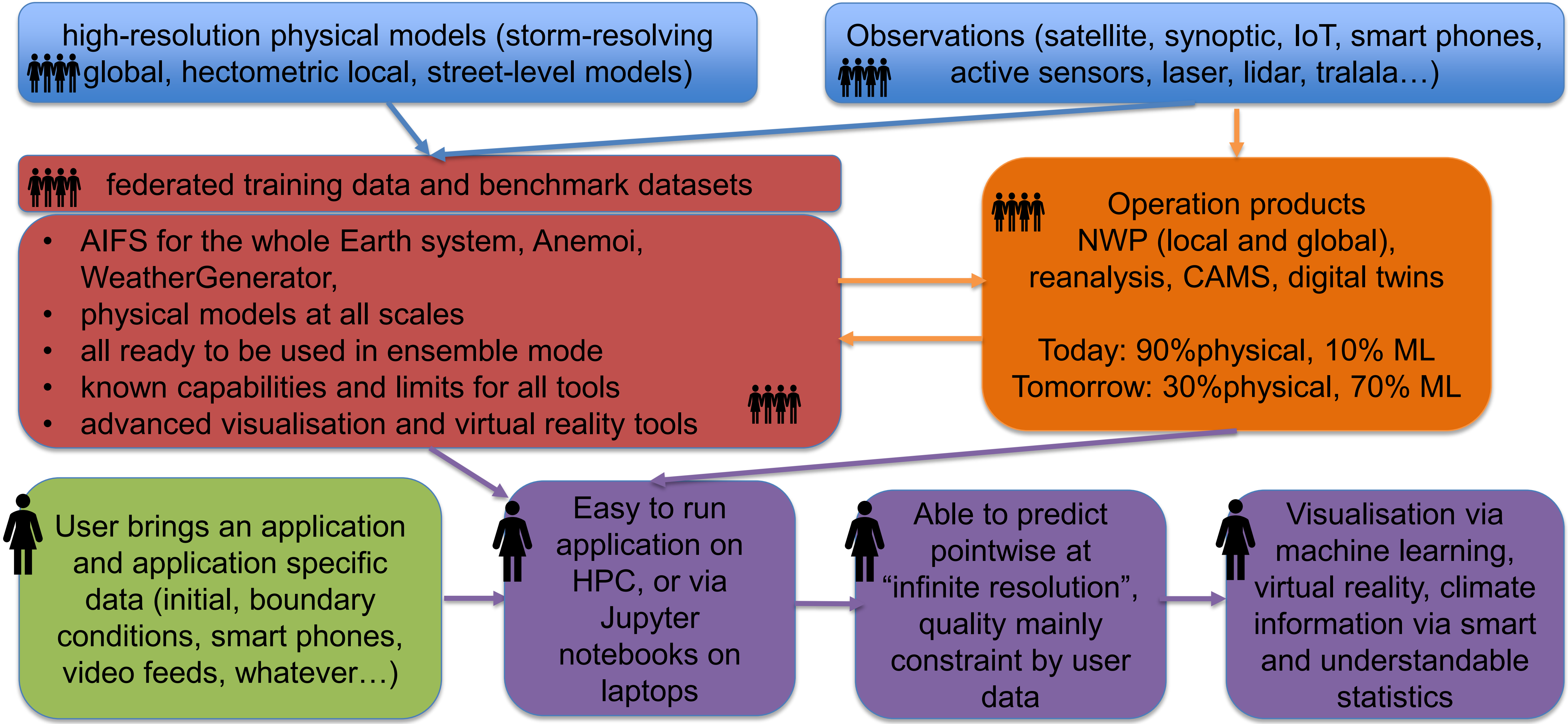
Let's call this a *common goal*.

The old hedgehog concept for Earth system modelling in the quiet revolution:
Build holistic physical models to allow for the best possible weather and climate prediction from days to seasons to climate.

The new hedgehog concept for Earth system modelling:
Use physical models and observations to create data to machine-learn models that provide the Earth system state at every single point in space and time including the past, present and future.



Future of Earth system models



ECMWF

Differences between physical and ML models, and modelling and services will blur. New user groups and customers can exploit the wealth of Earth science information.

Future use of Earth system models exploiting future AI

Distribution of NWP data: BBC radio → internet browser → mobile app → TikTok?

User can take a picture and scroll through the next two weeks and see in the picture how the weather is changing.

Users can also scroll through the decades of weather data from the 1950th to the 2050th and see how the climate zone is changing. Climate and uncertainty information is provided that includes probabilities of extreme events (such as precipitation, tropical cyclones, tornadoes, and heat waves).

Users take a video stream with their smart phone and virtually "drive" through their city at the day they were born, potentially combined with virtual reality and real-time video games.

Users trust social media tools to make suggestions for outdoor activities to the beach or the mountains taking weather and traffic predictions into account.

All tools are interactive via speech with LLM interface.

Many thanks! **Peter.Dueben@ecmwf.int**

Weather Generator

MAELSTROM

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Funded by the European Union

The WeatherGenerator project is funded by the European Union under grant agreement No 101187947. The MAELSTROM and ESiWACE projects have received funding from the EuroHPC-Joint Undertaking under grant agreement No 955513 and 101093054.